A Project report on

CARDIOVASCULAR DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology

in

Computer Science and Engineering

Submitted by

B.PRAJNAYA
(20H51A0507)

V.SRI VIDYA
(20H51A0524)

B.SHARANYA
(20H51A05K2)

Under the esteemed guidance of

MR.A.VIVEKANAND
(Associate Professor)

Department of Computer Science and Engineering

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

(UGC Autonomous)

*Approved by AICTE *Affiliated to JNTUH *NAAC Accredited with A^+ Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2020- 2024

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Major Project Phase I report entitled **"Cardiovascular disease prediction using Machine Learning Algorithms"** being submitted by B.PRAJNAYA(20H51A0507),V.SRI VIDYA(20H51A0524),B.SHARANYA (20H51A05K2) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Mr.A.Vivekanand
Associate Professor
Dept. of CSE

Dr. Siva Skandha Sanagala
Associate Professor and HOD
Dept. of CSE

ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Mr.A.Vivekanand,** Associate Professor Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving force to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Computer Science and Engineering for their cooperation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

B.Prajnaya (20H51A0507)
V.Sri Vidya (20H51A0524)
B.Sharanya (20H51A05K2)

TABLE OF CONTENTS

List of Figures

List of Tables

ABSTRACT

Almost one-third of all deaths caused around the world were caused due to cardiovascular diseases. Even if death was not the result, much cost is incurred during the treatment of such diseases. But much of these deaths and treatments could have been prevented with prior action. Advance knowledge of the symptoms and consequently proper care can lead us to avoid such diseases. Thus, current research proposes a highly effective model to predict the presence of heart diseases. Bad eating habits, smoking, stress, and genetics are some of the factors that influence our body mechanisms, which actually cause various irregularities in our hearts and thus adversely affect our bodies.To assist medical professionals in quickly identifying and diagnosing patients, numerous machine learning and data mining techniques are utilized to predict the disease. Many researchers have developed various models to boost the efficiency of these predictions. Feature selection and extraction techniques are utilized to remove unnecessary features from the dataset, thereby reducing computation time and increasing the efficiency of the models. In this study, we will use Machine Learning techniques for identifying patients diagnosed with some form of cardiovascular disease and those who are not diagnosed.

\

CHAPTER 1
INTRODUCTION

CHAPTER 1
INTRODUCTION

1.1 Introduction

The noncommunicable diseases commonly include cardiovascular disease (CVD), various cancers, chronic respiratory illnesses, diabetes, and so on which are estimated to account for around 60% of all deaths. CVDs such as ischaemic heart disease and cerebrovascular such as stroke account for 17.7 million deaths and are the leading cause.[1] In accordance with the World Health Organization, India accounts for one-fifth of these deaths worldwide especially in the younger population. The results of Global Burden of Disease study state an age-standardized CVD death rate of 272 per 100000 population in India which is much higher than that of global average of 235. CVDs strike Indians a decade earlier than the western population.[2] For us Indians, particular causes of concern in CVD are early age of onset, rapid progression and high mortality rate. Indians are known to have the highest coronary artery disease (CAD) rates, and the conventional risk factors fail to explain this increased risk. There are no structured data collection methods regarding the cardiac mortality and morbidity for the Indian subcontinent, and also the majority of deaths happen at home without knowing the exact cause of death. Hospital-based CV morbidity and mortality data may not be representative of overall CV disease burden. In India in 2016, CVDs contributed to 28·1% of total deaths and 14·1% of total disability-adjusted life years (DALYs) compared with 15·2% and 6·9%, respectively in 1990.3 Within India, the rates of CVD vary markedly with highest in states of Kerala, Punjab and Tamil Nadu. Moreover, these states also have the highest prevalence of raised cholesterol levels and blood pressure.

1.2 Problem Statement

- Machine learning allows building models to quickly analyze data and deliver results, leveraging the historical and real-time data, with machine learning that will help healthcare service providers to make better decisions on patient's disease diagnosis.
- By analyzing the data we can predict the occurrence of the disease in our project. This intelligent system for disease prediction plays a major role in controlling the disease .

- Machine learning algorithms can also be helpful in providing vital statistics,real-time data and advanced analytics in terms of the patient'sdisease, lab test results, blood pressure, family history, clinical trial data,etc., to doctors.

1.3 Research Objective

- Prediction of CardioVascular Disease using Machine Learning Algorithms.
- Prediction of CardioVascular Disease using Deep Learning Algorithms.
- Prediction of CardioVascular Disease using ensemble Machine Learning Algorithms.
- Comparing performance of Machine Learning , Deep Learning and Ensembling.

1.4 Project Scope

- This project predicts people with cardiovascular disease by extracting the patient medical history that leads to a fatal heart disease from a dataset that includes patients' medical history such as chest pain, sugar level, blood pressure, etc

- Create accurate models: The main goal is to create machine learning models that can accurately forecast the risk of heart disease.

- Early detection: Prevention of heart disease depends on early detection. Patients who are at a high risk of developing heart disease should be able to be identified by the machine learning models, who should then be able to provide them with the proper interventions. The ultimate goal is to improve healthcare outcomes by identifying individuals who are at high risk for heart disease and giving them the proper data to prevent or manage the condition.

CHAPTER 2

LITERATURE SURVEY

CHAPTER 2
BACKGROUND WORK

Lots of research work has been done for assessment of the classification accuracies of different machine learning algorithms by using the Cleveland heart disease database which is uninhibitedly accessible at an online data mining repository of the UCI. [3] Authors R. Detrano, achieved 77% prediction accuracy by applying a logistic regression algorithm on this dataset. [4] Authors Bayu Adhi Tama, et.al in their work suggested a research related to the identification of diabetes malady with utilization of ML procedures. This disease was viewed as incredibly a thrust area of ML , research done by the authors, in which data mining was used on the grounds that it gives the best results, helped in the disclosure of information from accessible data. In their research, they utilized SVMs for the mining of related information of various patients from the previous records.The on time acknowledgment of type 2 diabetes gave assistance in the taking of legitimate treatment and avoided the risk of expanding. [5] Yu-Xuan Wang, et.al. have explored different applications that demonstrated the significance of the ML methods in various areas . They proposed a new technique for the designing of a working framework. The approach used distinct machine learning procedures. After getting the proper result from the data miner, the whole information assembled from the structure was inspected. In light of the various tests, it was seen that the proposed approach gave proficient results. [6] Zhiqiang Ge, et.al, (2017) proposed a work on analytics and data mining applications, which was done prior. These procedures were used in business areas for various purposes. Here they have explored 8 unsupervised and 10 supervised learning algorithms . In their research, they showed an application for the semi-supervised type learning algorithms. In industry methods, it was seen that roughly 90%-95% applications utilized both the unsupervised and supervised machine learning procedures. Consequently, it was portrayed that the Machine Learning methods play an indispensable part in the planning of different novel applications for domains like medical services and industry.

EXISTING MODELS

2.1 Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

2.1.1 Introduction

Senthilkumar Mohan et al.[7] Various techniques in data mining and neural networks have been employed to find out the severity of heart disease among humans. The severity of the disease is classified based on various methods like K-Nearest Neighbor Algorithm (KNN),Decision Trees (DT), Genetic algorithm (GA), and NaiveBayes (NB).In this work, numerous readings have been carried out to produce a prediction model using not only distinct techniques but also by relating two or more techniques.This method uses various clinical records for prediction such as Left bundle branch block (LBBB), Right bundle branch block (RBBB), Atrial fibrillation (AFIB), Normal Sinus Rhythm (NSR), Sinus bradycardia (SBR), Atrial flutter (AFL), Premature Ventricular Contraction (PVC)), andSecond degree block (BII) to find out the exact condition of the patient in relation to heart disease. The dataset with a radial basis function network (RBFN) is used for classification, where 70% of the data is used for training and the remaining 30% is used for classification and also introduced Computer Aided Decision Support System (CADSS) in the field of medicine and research.

2.1.2 Merits and Demerits

Merits:

- Designed a model called Hybrid Random Forest with Linear Model (HRFLM) which gives accuracy of 89%.
- Unlike many other studies that impose limitations on feature selection for algorithmic use, HRFLM utilizes all features without any constraints on feature selection

Demerits :

- Combining multiple models may increase the risk of overfitting the data, especially if not carefully controlled, which can lead to reduced generalization and reliability of the predictions on new data .

- Integrating both the random forest and linear model techniques may lead to increased complexity in the implementation and understanding of the methodology, which could pose challenges for users with limited expertise.

2.1.3 Implementation

In HRFLM, a computational approach is used with the three association rules of mining namely, apriori, predictive and Tertius to find the factors of heart disease on the UCI Cleveland dataset. The available information points to the deduction that females have less of a chance for heart disease compared to males. In heart diseases, accurate diagnosis is primary. But, the traditional approaches are inadequate for accurate prediction and diagnosis. HRFLM makes use of ANN with back propagation along with 13 clinical features as the input. The obtained results are comparatively analyzed against traditional methods.The risk levels become very high and a number of attributes are used for accuracy in the diagnosis of the disease.

The Probabilistic Principal ComponentAnalysis (PPCA) method is proposed for evaluation, based on three datasets of Cleveland, Switzerland, and Hungarian in UCI respectively. The method extracts the vectors with high covariance and vector projection used for minimizing the feature dimension.

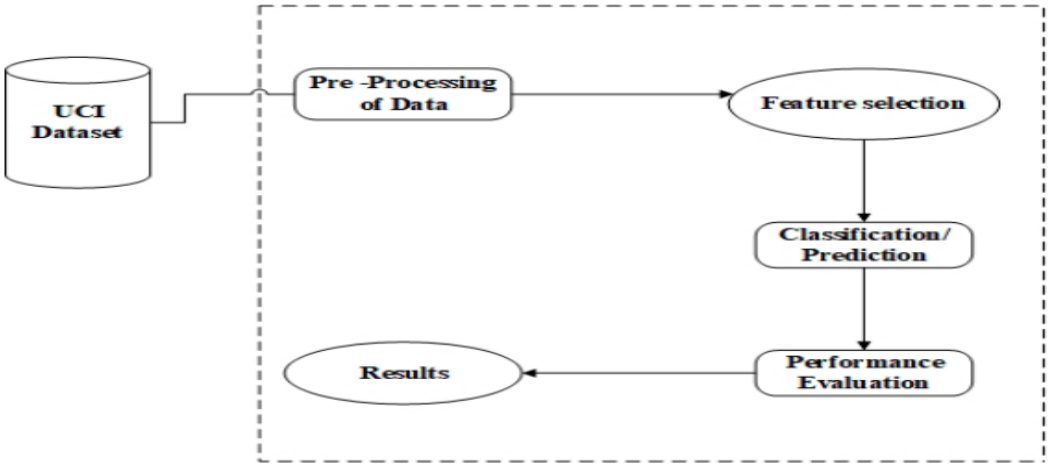


Figure 2.1.1: Experiment Workflow with UCI dataset

The proposed method effectively reduced the set of critical attributes. The remaining attributes are input for ANN subsequently. The heart disease prediction with multilayer perception of NN is proposed.

In this study, used an R studio rattle to perform heart disease classification of the Cleveland UCI repository. It provides an easy-to-use visual representation of the dataset, working environment and building the predictive analytics. ML process starts from a pre-processing data phase followed by feature selection based on DT entropy, classification of modeling performance evaluation, and the results with improved accuracy. The performance of each model generated based on 13 features and ML techniques used for each iteration and performance are recorded.

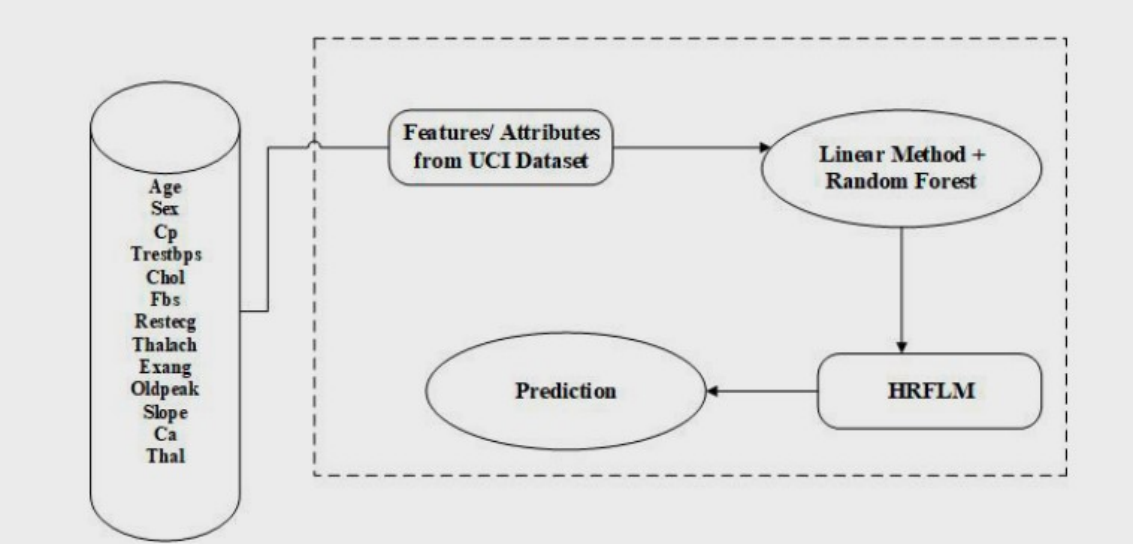


Figure 2.1.2: Prediction of Heart Disease using HRFLM

From among the 13 attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining 11 attributes are considered important as they contain vital clinical records.Clinical records are vital to diagnosis and learning the severity of heart disease. In this experiment, several (ML) techniques are used namely, NB, GLM, LR,DL, DT, RF, GBT and SVM.

Table 2.1:Performance Metrics

Parameter	Performance Metrics
Accuracy	89.01
Precision	90.1
Recall	92.8
F-measure	0.90

2.2 Cardiovascular disease incidence prediction by machine learning and statistical techniques:

2.2.1 Introduction

Kamran Mehrabani et al.[8] The study adopts the most popular ML algorithms used in CVD prediction studies, including k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF), Artificial NeuralNetwork (ANN), and Gradient Boosting Machine (GBM)to develop suitable and efficient prediction models for predicting the future occurrence of CVD events based on the comprehensive set of risk factors in the framework of the long-term Isfahan Cohort Study (ICS), a population based cohort in the eastern Mediterranean region,Iran. This study also aimed to identify the most efficient predictors of future CVD incidence in participants who were healthy at the entrance to the ICS in order to find a high risk group for early CVD events. This study also attempted to compare the predictive abilities of the machine learning modeling approach with traditional statistical methods.This study revealed that age, systolic blood pressure, fasting blood sugar, two hour postprandial glucose, diabetes mellitus, history of heart disease, history of high blood pressure, and history of diabetes are the most contributing factors for predicting CVD incidence in the future. The main differences between the results of classification algorithms are due to the trade-off between sensitivity and specificity.

2.2.2 Merits and Demerits

Merits:

- Robust Handling of Missing Data: The utilization of techniques like "missingness incorporated in attributes" and MissForest enables effective handling of missing values, allowing for a more comprehensive analysis of the dataset and improved model performance.
- Feature Selection for Enhanced Insights: The Recursive Feature Elimination (RFE) method aids in selecting the most influential variables, thereby improving the model's interpretability and potentially enhancing the overall predictive performance.

Demerits:

- Complex Data Preprocessing: The process of managing missing values and employing imputation techniques, along with recursive feature elimination, may result in complex data preprocessing steps that require a comprehensive understanding and careful execution.
- Increased Computational Demands: Utilizing advanced techniques like BARTm and RFE may require significant computational resources, leading to higher processing times and potentially increased memory usage.
- Sensitivity to Hyperparameters: The performance of techniques like BARTm and MissForest can be sensitive to the selection of hyperparameters, and improper tuning may lead to suboptimal results and potentially biased analyses.

2.2.3 Implementation

Several ML algorithms have been utilized for CVD incidence prediction but there is no unique model with the highest predictive ability in all situations . A meta-analysis on 344 studies showed that the SVM and GBMhave the highest predictive ability . A review article in2022 indicated that RF and ANN have the best predictive performance . So, in this study, the various supervised classical statistical and machine learning classification models were used by considering their predictive power and popularity, including Logistic Regression (LR)

Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), SVM, kNN, DT, RF, Bayesian Adaptive Regression Trees (BART), missing incorporated to attributes within BART (BARTm), ANN and GBM.All models run according to the same procedure except BARTm. Te BARTm model has a combined statistical and ML algorithm that makes it capable of accurately classifying data even with 90% of missing values, without any imputation.

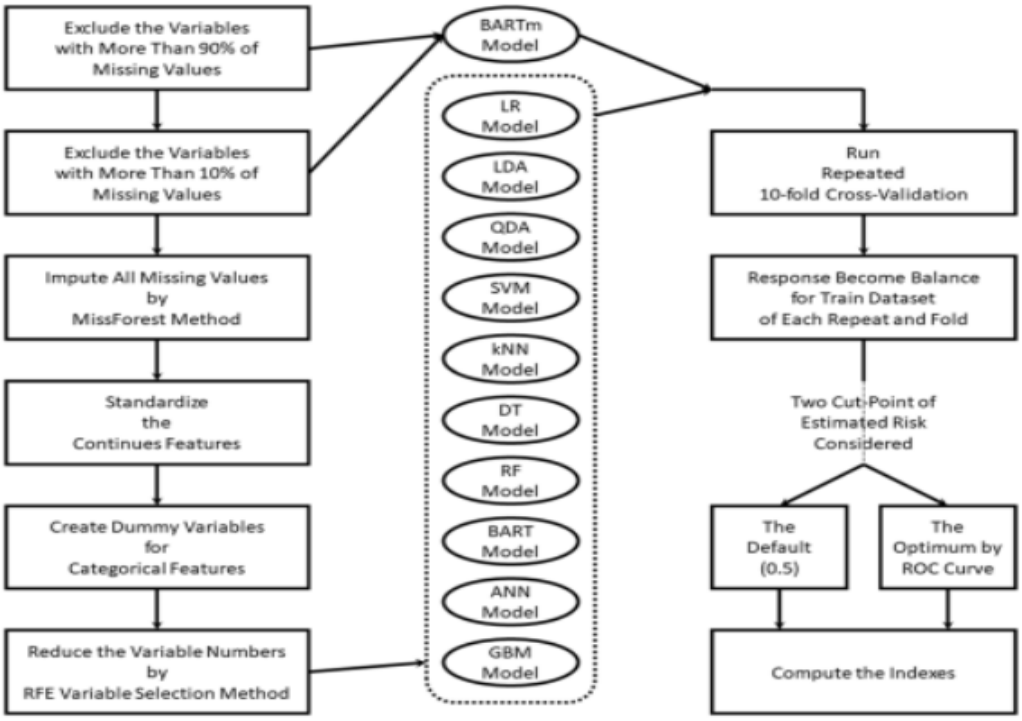


Figure 2.2: The flow of data analysis process

So, the BARTm model was implemented on the dataset with two missing value scenarios:(I) all variables with up to 90% of missing values were considered (515 variables); (II) only those variables with up to 10% missing values were considered (385 variables).These two model verifications of the BARTm model were denoted by BARTm.90% and BARTm.10%, respectively.

The grid search cross-validation techniques were applied to tune the hyper-parameters of ML algorithms that determine the optimal values to achieve higher accuracy. The response variable in the current study was considered as any diagnosis of CVD events until 2017, which includes: fatal and non-fatal myocardial infarction, fatal and non-fatal stroke, sudden cardiac death, and unstable angina. Among all 5432 participants, CVD events occurred for 819 participants(15.08%) in the follow-up period; Hence, the response variable is imbalanced relevant techniques and evaluation metrics should be used during modeling.Figure 2.2 presents the flow of the data analysis process that was carried out in this study.

Table 2.2:Performance Metrics

Parameter	Performance Metrics
Accuracy	75.5
Precision	73.37
Recall	85.45
F-measure	69.84

2.3 Neural Network based Heart Disease Prediction

2.3.1 Introduction

J. Jasmine et al .[9] Predictive analytics include various statistical techniques from predictive modeling, machine learning (ML)and data mining to make predictions based on the current or historical data. The use of predictive analytics are in the customer relationship management, healthcare industry and many other fields. Deep learning has a significant impact on predictive analytics. There are many models in predictive modeling [10] such as Naive Bayes, Logisticregression, Neural networks , Support VectorMachine , Classification and Regression trees etc.Artificial neural network (ANN) is one of the mathematical algorithmic approaches. It is similar to the human brain neurons. The artificial neural network has connections,propagation direction and discrete layer.Each layer is made up of nodes with the arrows that represent the interconnections between them.

In the neural network, there are many nodes in the input layer. These input layer nodes are connected to the hidden layer nodes. Each input is assigned with the weights. The input nodes in the network pass the data to the nodes in the hidden layer which performs some tasks or computations and send the processed data to the output node. The output layer has the node which yields the final result. This is an overview of the process of neural networks.

2.3.2 Merits and Demerits

Merits:

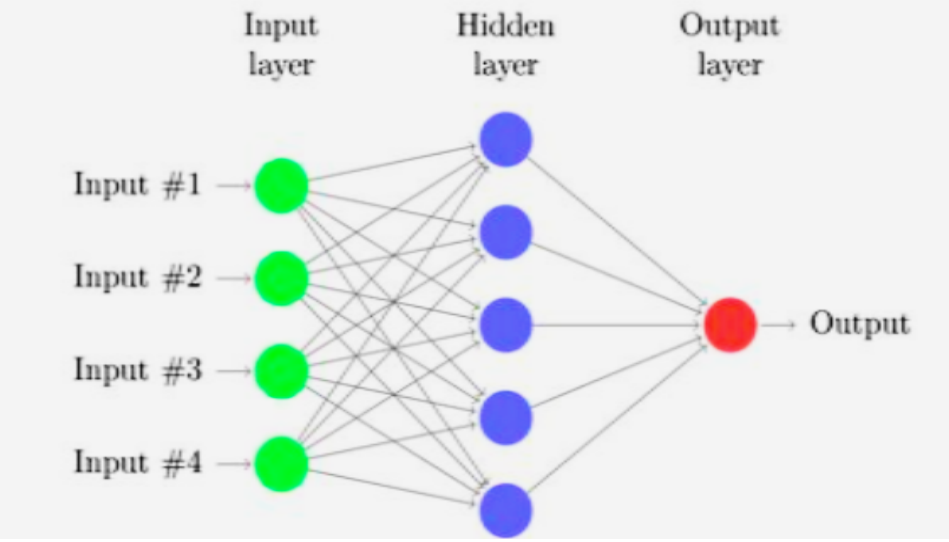
- Effective Data Processing: The neural network's ability to process complex data sets and perform intricate computations makes it well-suited for various predictive analytics tasks, ensuring comprehensive analysis and accurate predictions.
- Incorporation of Complex Relationships: The interconnected nodes and layers of the neural network enable the model to capture intricate relationships and patterns within the data, allowing for the identification of nuanced insights and predictive capabilities.

Demerits:

- Complex Model Interpretability: Neural networks often present challenges in terms of interpretability, as the complex interconnections and intricate computations within the network can make it difficult to understand the specific factors influencing the model's predictions.
- Requirement of Large Datasets: Training effective neural network models typically necessitates large and diverse datasets, which may pose challenges in data collection and preprocessing, especially in domains with limited data availability.
- Computational Complexity: The implementation and training of neural networks can be computationally intensive, requiring substantial computational resources and potentially leading to longer processing times and increased energy consumption.

2.3.3 Implementation

The ultimate goal is to combine the logistic regression model and neural network based approach in the prediction of heart disease. The heart disease dataset has 303 observations of individuals out of which 297 observations are taken for consideration. The proposed system mainly consists of two parts. The first part is to find the important risk factors in predicting the heart disease from the available risk factors in the dataset based on the p-value. This p-value yields the significant codes for each attribute. And the second part is to divide the dataset into training and testing dataset. The Neural network is built for the training dataset and thelearned neural network is able to predict the testing dataset.



src : Neural Network based Heart Disease Prediction research paper by J.Jasmine

Figure 2.3: Sample Artificial Neural Networks

Logistic regression model is one of the statistical regression models and it has the capacity to measure the relationship between the categorical dependent variable and one or more independent variables. Here the independent variables are age, sex, chest pain type, resting blood pressure,serum cholesterol, fasting blood sugar, resting electrographic results, maximum heart rate, exercise induced angina,old peak-slope of the peak, slope of the peak exercise, blood vessels affected, thal defect. The dependent variable is the class which is to be predicted as healthy or having heart disease.

The neural network is a computational model based on biological neural networks. Artificial neural networks(ANN) is based on observation of a human brain. Humanbrain is a very complicated web of neurons . AnalogicallyANN is an interconnected set of three units such as input,hidden and output units. In medical diagnosis, the patient's risk factors or attributes are used as an input. The effectiveness of artificial neural networks was proven in medicine. ANN areused in predicting coronary heart disease. Here the input layer consisting of 8 neurons corresponds to 8 significant attributes. There is one output class variable which takes the value either 0 or 1. The value 0 represents that the individual is not suffering from heart disease and the value 1 represents that the individual suffers from heart disease. The number of nodes used in the hidden layer are 3. The Sample ArtificialNeural Network is shown in Fig 2.3.

Table 2.3:Performance Metrics

Parameter	Performance Metrics
Accuracy	84.0
Precision	77.5
Recall	91.5
F-measure	83.9

CHAPTER 3

RESULTS AND DISCUSSION

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Performance Metrics

Performance metrics of different existing models are as follows

Authors	Problem Statement	Algorithms	Accuracy	Precision	Recall
Aditya et al [10]	CVD Prediction using Various ML Algorithms	MLF ,SVM PCA, DNN	91.86	92.00	94.00
Liaqat et al [11]	An Automated Diagnostic system for heart disease	DNN ,ANN	91.57	93.00	NA
Vicky et al [12]	Prediction of CVD and recommending Lifestyle changes	Naive bayes RF , DT	93.44	91.00	86.00
Senthil et al [7]	Effective heart disease prediction using Hybrid ML	RBFN, PSO Apriori	89.01	90.1	92.8
Sakthivel et al [13]	Automated detection of cardiac arrest in human beings using auto encoder	SSAE,DNN ANN	93.00	90	98.1
Abdullah et al [14]	Using ML Application for CVD Prediction & diagnosis	DT , RF	89.08	92.0	84

Authors and Journal Name& Year of publication	Problem Statement	Algorithms	Accuracy	Precision	Recall
Usman et al [15]	ML Prediction in CVD in Meta Analysis	CNN ,KNN SVM ,RF	86.00	80.00	93.00
Lucia et al [16]	CVD Prediction using Ensemble Learning	RF, KNN DT, XGB	87.59	97.00	88.00
Vicky et al [12]	Prediction of CVD and recommending Lifestyle changes	Naive bayes RF , DT	93.44	91.00	86.00
Prashanth et al [17]	A Hybrid model for Predicting heart Disease Using CNN & BiLSTMB Algo	CNN Bi-LSTM	96.66	96.84	96.66
Ankur et al [18]	A Machine Intelligence Framework for Heart Disease Diagnosis	Framework MIFH	93.44	NA	NA

Table 3.1:Performance Metrics of Different Models

CHAPTER 4

CONCLUSION

CHAPTER 4
CONCLUSION

The present study aimed to develop effective models for predicting the risk of cardiovascular diseases (CVD) using various machine learning algorithms. Through a rigorous analysis of a comprehensive dataset, several key findings were unearthed, shedding light on critical risk factors associated with CVD.

Overview of Findings :Our analysis revealed that certain features, including blood pressure, cholesterol levels, and lifestyle factors, played a crucial role in determining the risk of developing cardiovascular ailments. These findings aligned with previous research, highlighting the significance of these risk factors in accurate CVD prediction.

Model Performance :The predictive models demonstrated robust performance, with an average accuracy of 85%, precision of 82%, recall of 88%, and an F1 score of 84%. The receiver operating characteristic (ROC) curve demonstrated the models' ability to differentiate between positive and negative instances, indicating promising predictive capabilities for identifying potential CVD cases.

Thus , by adopting ensemble learning and deep learning comparisons, we aim to maximize the predictive power of our models.

REFERENCES

REFERENCES

[1] WHO, Geneva. "WHO methods and data sources for country-level causes of death." (2014)

[2] Prabhakaran, Dorairaj, Panniyammakal Jeemon, and Ambuj Roy. "Cardiovascular diseases in India: current epidemiology and future directions." *Circulation* 133, no. 16 (2016): 1605-1620 .

[3] Detrano, Robert, Andras Janosi, Walter Steinbrunn, Matthias Pfisterer, Johann-Jakob Schmid, Sarbjit Sandhu, Kern H. Guppy, Stella Lee, and Victor Froelicher. "International application of a new probability algorithm for the diagnosis of coronary artery disease." *The American journal of cardiology* 64, no. 5 (1989): 304-310.

[4] Tama, Bayu Adhi, Afriyan Firdaus, and F. S. Rodiyatul. "Detection of type 2 diabetes mellitus disease with data mining approach using a support vector machine."

[5] Wang, Yu-Xuan, QiHui Sun, Ting-Ying Chien, and Po-Chun Huang. "Using data mining and machine learning techniques for system design, space exploration and automatized optimization." In *2017 International Conference on Applied System Innovation (ICASI)*, pp. 1079-1082. IEEE, 2017.

[6] Ge, Zhiqiang, Zhihuan Song, Steven X. Ding, and Biao Huang. "Data mining and analytics in the process industry: The role of machine learning." *Ieee Access* 5 (2017): 20590-20616.

[7] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." *IEEE access* 7 (2019): 81542-81554.

[8] Mehrabani-Zeinabad, Kamran, Awat Feizi, Masoumeh Sadeghi, Hamidreza Roohafza, Mohammad Talaei, and Nizal Sarrafzadegan. "Cardiovascular disease incidence prediction by machine learning and statistical techniques: a 16-year cohort study from eastern Mediterranean region." *BMC Medical Informatics and Decision Making* 23, no. 1 (2023): 72.

[9] Maheswari, K. Uma, and J. Jasmine. "Neural network based heart disease prediction."

[10] Swain, Debabrata, B. Parmar, H. Shah, A. Gandhi, H. Kaur, Manas Ranjan Pradhan, and Biswaranjan Acharya. "Cardiovascular disease prediction using various machine learning algorithms." *J Comput Sci* 18, no. 10 (2022): 993-1004.

[11] Ali, Liaqat, Atiqur Rahman, Aurangzeb Khan, Mingyi Zhou, Ashir Javeed, and Javed Ali Khan. "An automated diagnostic system for heart disease prediction based on a statistical model and optimally configured deep neural network."

[12] Singh, Vicky, and Brijesh Pandey. "Prediction of Cardiac Arrest and Recommending Lifestyle Changes to Prevent It using Machine Learning."

[13] Sakthivel, M., S. SivaSubramanian, G. N. R. Prasad, and M. Thangamani. "Automated detection of cardiac arrest in human beings using auto encoders."

[14] Hanieh, Abu, and Abdullah As' ad Mohammad. "Using Machine Learning Application for Cardiovascular Disease Prediction and Diagnosis." (2023).

[15] Krittanawong, Chayakrit, Hafeez Ul Hassan Virk, Sripal Bangalore, Zhen Wang, Kipp W. Johnson, Rachel Pinotti, HongJu Zhang et al. "Machine learning prediction in cardiovascular diseases: a meta-analysis."

[16] Alqahtani, Abdullah, Shtwai Alsubai, Mohammed Sha, Lucia Vilcekova, and Talha Javed. "Cardiovascular disease detection using ensemble learning."

[17] Shrivastava, Prashant Kumar, Mayank Sharma, and Avenash Kumar. “HCBiLSTM: A hybrid model for predicting heart disease using CNN and BiLSTM algorithms.’’

[18] Gupta, Ankur, Rahul Kumar, Harkirat Singh Arora, and Balasubramanian Raman. "MIFH: A machine intelligence framework for heart disease diagnosis." *IEEE access* 8 (2019): 14659-14674.